

Testing To, and Beyond, the Common Core

By Linda Darling-Hammond

New assessments can support a multiple-measure framework to deepen teaching and learning.



After more than a decade of test-driven, high-stakes accountability in the No Child Left Behind era, many educators and policymakers in the United States are looking to move toward a more thoughtful approach. Rather than maintaining a system that uses narrow measures of student achievement to sanction poorly performing schools, the push is now to implement next-generation learning goals that encourage higher-order thinking skills.

The driving force behind this shift is the Common Core State Standards (CCSS) for English language arts and mathematics. Forty-five states and the District of Columbia have adopted the standards. State-led initiatives—such as the Next Generation Science Standards and Common Career Technical Core—are next in line.

A critical piece in this roadmap will be new assessments, which have the potential to give school leaders new and better tools to guide instruction, support teachers, and improve outcomes. Assessment decisions will have a big impact on principals, who know

Principal ONLINE

Access the following Web resources by visiting *Principal* magazine online: www.naesp.org/JanFeb14

✦ “Implementing the Common Core State Standards,” a joint action brief from NAESP, outlines **the role of elementary school principals** in CCSS implementation.

✦ NAESP’s Common Core Implementation Checklist is an **easy-to-use worksheet** school leaders can use to track progress with CCSS.

✦ The Jan/Feb 2012 issue of *Principal* magazine focused on aspects of **Common Core implementation** such as considering students with disabilities, English language arts standards, and using technology.

the difference between leading a school constrained by punitively used tests that fail to measure many of the most important learning goals, and a school that uses thoughtful assessments to measure what matters and inform instruction.

If we are to achieve 21st century standards for learning, it is critical that these new assessments:

- **Are much broader than the standardized tests of the No Child Left Behind (NCLB) era.** They must measure the full range of higher-order thinking skills and important education outcomes, including critical thinking, communication, collaboration, social-emotional competence, moral responsibility, and citizenship.
- **Are part of a framework that considers multiple measures of valued outcomes in all decisions about students, educators, and schools.** As advised by the Psychological Standards on Testing, decisions about student promotion, placement, and graduation—as well as teacher, principal, and school evaluation—should never be based on a single test, but on a combination of classroom and school measures appropriate to the students, curriculum, and context of the decision.
- **Become part of a new accountability system that replaces the old test-and-punish philosophy with one**

that aims to assess, support, and improve. Tests should be used not to allocate sanctions, but to provide information, in conjunction with other indicators, to guide educational improvement.

Moving Beyond NCLB

When it comes to student testing in the United States, it is clear that changes are needed. The public doesn’t trust most tests in use today. In the *2013 PDK/Gallup Poll of the Public’s Attitudes Toward the Public Schools*, only 22 percent of respondents said increased testing had helped the performance of their local schools, a decrease from 28 percent in 2007. More striking, 36 percent of those questioned said the testing was hurting school performance; 41 percent said it had made no difference.

Educators are also increasingly leery of current assessments and how they are used. Last year, *Primary Sources: 2012*, a report by Scholastic and the Gates Foundation, found that only 28 percent of educators see state-required standardized tests as an important gauge of student achievement. In addition, only 26 percent of teachers say standardized tests are an accurate reflection of what students know.

This collective skepticism is a reaction to a decade of tests that almost exclusively emphasize low-level skills. A growing number of parents and educators are uncomfortable with the fact that today’s students are drilling for multiple-choice tests geared to the expectations of the past. It’s in this context that CCSS offer an opportunity to pivot toward a richer and more rigorous system of assessment.

An Opportunity to Improve Assessment Systems

Because the CCSS are intended to be “fewer, higher, and deeper” than previous standards, they have created a natural opening for the development and adoption of better assessments of student learning. The assessments

developed by two new multi-state consortia could move us toward more informative systems that include formative as well as summative elements, evaluate content that reflects instruction, and include some challenging open-ended tasks.

These assessments, though, will not include all necessary tasks and skills for students, such as long-term research and investigation tasks or the ability to communicate orally, visually, and with technology tools. These kinds of tasks are needed to develop and assess students’ abilities to find and use information to solve problems, explain different approaches to a problem, and explain and defend their reasoning. That is why some schools, districts, and states are developing more robust performance tasks and portfolios as part of multiple-measure systems of assessment. In addition to CCSS-aligned consortia exams, multiple measures could include:

- Classroom-administered performance tasks (e.g., research papers, science investigations, mathematical solutions, engineering designs, arts performances);
- Portfolios of writing samples, art works, or other learning products;
- Oral presentations and scored discussions; and
- Teacher rating of student note-taking skills, collaboration skills, persistence with challenging tasks, and other evidence of learning skills.

These activities not only engage students in more intellectually challenging work that reflects 21st century skills, they also serve as learning opportunities for teachers, when they are involved in using the assessments



and scoring them together. Priti Johari, the redesign administrator for Chelsea High School in Massachusetts notes about her school’s efforts:

Our work of creating common performance assessments and rubrics and scoring them across classrooms has created a culture of inquiry and a collaborative atmosphere... This is a result of our process of learning about the Common Core, unpacking standards, writing lesson plans and tasks, sharing those plans, giving each other feedback, creating common rubrics, and collectively examining student work.

Two decades of research has found that when teachers use, score, and discuss the results of high-quality performance assessments over time, both teaching and learning improve. Teachers become expert in their practice and more attuned to how students think and learn. Meanwhile, students learn to internalize standards and improve their own work, as they work on tasks guided by rubrics against which they self-assess and are assessed by peers and teachers.

In New Hampshire, where the new accountability system will rely substantially on a bank of complex performance tasks developed and scored by teachers with support from the state, deputy commissioner Paul Leather explains, “We want to move forward on a continuum toward deeper assessment that is more challenging for students and teachers. We are aiming eventually to have a system where the students create their own tasks and teachers score them with common rubrics.”

If used wisely, performance assessments have the potential to address multiple important education goals through one concerted investment. Not only will pedagogical capacity be

enhanced, but assessment will remain focused on its central purpose: the support of learning for all involved.

Supporting Better Teaching

In addition to supporting professional development, high-quality performance assessments can be part of a basket of evidence about student learning for teacher evaluation. Assessments that provide direct evidence of what students can do related to the specific curriculum they are taught can be more accurate and productive than the value-added metrics based on state test scores that are currently popular.

Although the idea of measuring teachers’ contributions to student learning through gains on standardized tests is appealing—and has been very valuable for large-scale studies—it turns out that, at the individual teacher level, value-added models (VAM) have many pitfalls. These are particularly problematic when state tests are used.

In addition to the fact that the tests are narrow and do not measure higher-order thinking skills, researchers have found that value-added models of teacher effectiveness are highly unstable: Teachers’ ratings differ substantially from class to class and from year to year, as well as from one test to the next. This is in part because there are many other influences on student gains other than individual teachers, and in part because teachers’ value-added ratings are affected by differences in the students who are assigned to them, even when statistical models try to control for student demographic variables.

In particular, teachers with large numbers of new English learners and other students with special needs have been found to show lower gains than the same teachers when they are teaching other students. This, in turn, is partly because—due to rules under NCLB—state tests are designed to measure only grade-level standards, which means they cannot assess growth for

Two decades of research has found that when teachers use, score, and discuss the results of high-quality performance assessments over time, both teaching and learning improve.

students who are either below or above grade-level, since there are no questions on the tests that are designed to measure that content.

As a result, VAM results can be extremely inaccurate for teachers. Consider, for example, the case of Carolyn Abbott. Ms. Abbott was a seventh- and eighth-grade math teacher at a New York school for gifted students. Beloved by students and parents alike, in 2010 her seventh graders scored at the 98th percentile on the city math test, many already hitting the top score (and thus unable to show growth). When she had these same students in eighth grade the next year, where they mostly worked on high-school level material, all of them passed the tenth-grade Regents test and fully one-third had perfect scores.

There was a problem, though: Although they did extremely well, Ms. Abbott's students hadn't shown "growth" on the eighth-grade state test, because it could not measure what they had learned beyond the grade level. This fact led to her being ranked as the worst eighth-grade math teacher in New York City on the value-added metric. Although her principal thought she was a great teacher and wanted her to stay, the rules for tenure in New York stood in the way. Ms. Abbott left to enter a Ph.D. program, and public education lost a great teacher.

This case is not unusual. A minority of teachers actually teach classrooms of students who are all achieving at grade level. The solution is to develop a basket of evidence about student learning gains that is appropriate for the curriculum and the students being taught. In Ms. Abbott's case, for example, her

basket of evidence might have included the tenth grade Regents test, perhaps with a pre-test she had designed to evaluate student needs at the beginning of the year and growth by the end of the year. It could also have included pre- and post-tests from a particular unit she wanted to focus on improving, and evidence from students' interdisciplinary math/science projects, designed to allow them to apply mathematics in a real world context. This would inform the entire teacher team working on the projects together.

A number of states and districts have devised multiple-measures approaches to teacher evaluation that combine classroom observations with a basket of evidence about student learning, as well as evidence about professional contributions. Sometimes, teacher teams work on their targets and strategies together, enhancing collaboration more powerfully. These kinds of systems also improve student learning, as teachers set goals on meaningful targets that they track using authentic evidence that emerges directly from classroom work. I discuss these systems in my 2013 book, *Getting Teacher Evaluation Right: What Really Matters for Effectiveness and Improvement*.


Toward Assessments That Improve Learning

Assessment can be, and should be, instructive for educators. A 21st century education system has no place for the antiquated distinction between teaching and testing. Modern assessments should provide valuable information to educators on their practice as well as insights about how individual students are doing.

In the coming years, principals will

have a chance to help construct systems of assessment that help improve learning—for teachers, parents, students, and policymakers. Questions that principals might ask themselves in this new era include:

- How can we engage students in assessments that measure higher-order thinking and performance skills—and use these to transform practice?
- How can these assessments be used to help students become independent learners, and help teachers learn about how their students learn?
- How can teachers be enabled to collect evidence of student learning that captures the most important goals they are pursuing, and then to analyze and reflect on this evidence—individually and collectively—to continually improve their teaching?
- What is the range of measures we believe could capture the educational goals we care about in our school? How could we use these to illustrate and extend our progress and successes as a school?

For principals, the new focus on high-quality assessment represents a critical juncture. As instructional leaders and catalysts for change, principals can work with teachers to develop, select, and use more productive assessment options that can help improve instruction and guide school improvement. 

Linda Darling-Hammond is the Charles E. Ducommun Professor of Education at the Stanford Graduate School of Education.